MSBD 5007 Homework 2

Pranav A, 20478966

October 17, 2017

Problem 1.

For the given $f(\mathbf{x})$, we know that sum of convex functions is a convex function. Hence we will prove that $l(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)$ is a convex function using Hessian.

For a sigmoid function $g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)$, we have,

$$\nabla_x g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i) = \nabla_x \left(\frac{1}{1+e^{-\mathbf{a}_i^T\mathbf{x}\mathbf{b}_i}} \right)$$
$$\nabla_x g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i) = \mathbf{b}_i\mathbf{a}_i \frac{e^{-\mathbf{a}_i^T\mathbf{x}\mathbf{b}_i}}{\left(1+e^{-\mathbf{a}_i^T\mathbf{x}\mathbf{b}_i}\right)^2}$$
$$\nabla_x g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i) = \mathbf{b}_i\mathbf{a}_i g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)\left(1 - g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)\right)$$

For the given loss function $l(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)$ leaving the outer constant out, the gradient is given by,

$$\begin{aligned}
&\nabla_x l(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i) \\
&= \nabla_x \ln\left(1+e^{-\mathbf{a}_i^T\mathbf{x}\mathbf{b}_i}\right) \\
&= \nabla_x - \ln g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i) \\
&= -\frac{1}{g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)} \nabla_x g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i) \\
&= -\frac{1}{g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)} \mathbf{b}_i\mathbf{a}_i g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)\left(1 - g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)\right) \\
&= \mathbf{b}_i\mathbf{a}_i\left(g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i) - 1\right)
\end{aligned}$$

The hessian of the $l(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)$ is given by,

$$\begin{aligned}
&\nabla_x^2 l(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i) \\
&= \nabla_x \nabla_x l(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i) \\
&= \nabla_x \left(\mathbf{b}_i\mathbf{a}_i\left(g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i) - 1\right)\right) \\
&= \mathbf{b}_i\mathbf{a}_i \nabla_x g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i) \\
&= \mathbf{b}_i\mathbf{a}_i\mathbf{a}_i^T\mathbf{b}_i g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)\left(1 - g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)\right) \\
&= \mathbf{b}_i^2\mathbf{a}_i\mathbf{a}_i^T\mathbf{b}_i g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)\left(1 - g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)\right)
\end{aligned}$$

To show that $l(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)$ is convex, we need to show that hessian is positive semi-definite, that means, all entries in $z^T\nabla_x^2 l(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)z$ should be greater than or equal to zero, for all vectors of z. Thus, we have,

$$\begin{aligned}
&\forall z, z^T\nabla_x^2 l(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)z \\
&= z^T\mathbf{b}_i^2\mathbf{a}_i^T\mathbf{a}_i\mathbf{b}_i g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)\left(1 - g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)\right)z \\
&= g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)\left(1 - g(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i)\right)z^T z\mathbf{b}_i^2\mathbf{a}_i^T\mathbf{a}_i
\end{aligned}$$

The terms $z^T z\mathbf{b}_i^2\mathbf{a}_i^T\mathbf{a}_i$ are positive and sigmoidal functions are positive. Therefore, entries of hessian matrix are equal or greater than zero. Thus the given loss function is convex. And the final sum, would be a convex function too.

Problem 2.1.

For the given $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{b}$, we would find the hessian of the given equation. Thus,

$$\begin{aligned}
\nabla_x f(\mathbf{x}) &= \nabla_x\left(\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{b}\right) \\
&= \frac{1}{2}\nabla_x\mathbf{x}^T\mathbf{A}\mathbf{x} - \nabla_x\mathbf{x}^T\mathbf{b} \\
&= \frac{1}{2}\left(A^T + A\right)\mathbf{x} - \mathbf{b} \\
&= \mathbf{A}\mathbf{x} - \mathbf{b} \quad \text{(becuase } A \text{ is symmetric)} \\
\nabla_x^2 f(\mathbf{x}) &= \nabla_x\nabla_x f(\mathbf{x}) \\
&= \nabla_x\left(\mathbf{A}\mathbf{x} - \mathbf{b}\right) \\
&= \mathbf{A}
\end{aligned}$$

Since A is symmetric positive semi definite, that means its Hessian is symmetric positive semi definite. This means that given function is convex.

Problem 2.2.

The formula of steepest descent is given by,

$$x^{k+1} = x^k - \alpha\nabla_x f(\mathbf{x})$$

where α is the step size and $\nabla_x f(\mathbf{x})$ is the gradient which was $Ax - b$.

Thus the formula of steepest descent looks like,

$$x^{k+1} = x^k - \alpha(Ax - b)$$

For the exact line search, the value of the step size is available, which is given by,

$$\alpha = argmin_\alpha f(x^k - \alpha\nabla_x f(\mathbf{x^k}))$$

Thus $f(x^k - \alpha\nabla_x f(\mathbf{x^k}))$ is given by,

$$\frac{1}{2}(\mathbf{x^k} - \alpha\nabla_\mathbf{x}\mathbf{f}(\mathbf{x^k})^T)\mathbf{A}(\mathbf{x^k} - \alpha\nabla_\mathbf{x}\mathbf{f}(\mathbf{x^k})) - (\mathbf{x^k} - \alpha\nabla_\mathbf{x}\mathbf{f}(\mathbf{x^k})^T)\mathbf{b}$$

Here we will get a quadratic equation with respect to α after substituting $\nabla_x f(\mathbf{x})$.

$$\frac{\alpha^2}{2}\left(x^TA^TAx - x^TA^TAb\right) - \alpha\left(x^TA^TAx - b^Tb - \frac{x^TAb}{2}\right) + c$$

Solving the minimum for that α (by differentiating above function with respect to α, and equating to 0) we get,

$$\alpha = \frac{x^TA^TAx - b^Tb - \frac{x^TAb}{2}}{x^TA^TAx - x^TA^TAb}$$

Problem 3.1.

We show that any norm $||\cdot|| : \mathbb{R}^n \to \mathbb{R}_+$ is a convex function.

It's clear the domain $\mathbb{R}^n$ is a convex set. Then by properties of the norm, $||kx|| = |k|||x||$ and $||x+y|| \le ||x|| + ||y||$, for any $x, y \in \mathbb{R}^n$ and $0 \le \theta \le 1$,

$$||\theta x + (1-\theta)y|| \le ||\theta x|| + ||(1-\theta)y|| = \theta||x|| + (1-\theta)||y||$$

Thus from the above definition, $||x||_2$ is a convex function.

Problem 3.2.

We know that $||x||_2 = \sqrt{x^Tx}$, if $x \ne 0$. Thus $\nabla||x||_2$ is given by,

$$\begin{aligned}\nabla f(x) &= \nabla||x||_2 \\ &= \nabla\sqrt{x^Tx} \\ &= \frac{1}{2}(x^Tx)^{\frac{-1}{2}}\nabla(x^Tx) \\ &= \frac{1}{2}\frac{2x}{\sqrt{x^Tx}} \\ &= \frac{x}{||x||_2}\end{aligned}$$

Problem 3.3.

We will use the definition of subdifferential here. We know that subdifferential $\partial f(x)$ of f at x is the set of subgradients, $\partial f(x) = \{g \,|\, g^T(y-x) \le f(y) - f(x), \forall y \in domain(f)\}$.

Thus it's easy to see that $\partial f(x)$ at $x = 0$ would be, $\partial f(x) = \{g \,|\, ||g||_2 \le 1\}$. This is true because supremum of $\partial f(x)$ is 1 and thus the subdifferential at $x = 0$ will not exceed the norm of 1.

Notes

1. This given solutions of the assignment follows the HKUST honour code. Although assignment has been discussed with other peers, the solutions are my own.

2. Kindly give feedback on this assignment on how to write up the solutions more elegantly. I apologize my omission to detailed steps due to my prematurity to $\LaTeX$.